

## Pergunta 1

Correta

Pontuou 1,0 de 1,0

🚩 Destacar pergunta

Which of these is not a predictive task:

- ☐ a. Não quero responder
- ☐ b. Recommendation systems.
- ☐ c. Regression.
- ☐ d. Rating.
- ☒ e. Clustering.



A resposta correta é: Clustering.

## Pergunta 2

Correta

Pontuou 1,0 de 1,0

🚩 Destacar pergunta

Choose from the options below the one that represents a data mining task:

- ☐ a. Extract the frequency from an electromagnetic wave.
- ☒ b. Predict sales volume based on historical data.
- ☐ c. Predict the face of a die for a new roll.
- ☐ d. Não quero responder
- ☐ e. Calculate the total sales volume for a given product.



A resposta correta é: Predict sales volume based on historical data.

### Pergunta 3

Correta

Pontuação 1,0 de 1,0

🚩 Destacar pergunta

The CRISP-DM data preparation phase is essential to the success of a data mining project because it:

- ☐ a. Não quero responder
- ☐ b. consists of a set of operations that, when applied correctly, guarantees the extraction of useful knowledge by the algorithms.
- ☐ c. consists of a set of tasks for robust evaluation of the models developed in previous phases.
- ☐ d. identifies problems with the data, which, if not identified, would lead to the creation of an unsatisfactory solution.
- ☒ e. aims to clean and transform the data in order to facilitate the extraction of knowledge by the algorithms.



A resposta correta é: aims to clean and transform the data in order to facilitate the extraction of knowledge by the algorithms.

### Pergunta 4

Correta

Pontuação 1,0 de 1,0

🚩 Destacar pergunta

Outlier detection techniques are used to identify observations with values

- ☐ a. of the target variable that are very different from what is common, and which are those that are to be identified in prediction problems.
- ☐ b. of the target variable that are very different from what is common, and that may hinder the detection of patterns.
- ☒ c. of the variables that are very different from what is common, and that can hinder the detection of patterns.
- ☐ d. Não quero responder
- ☐ e. of the variables that are very different from what is common, which are those that are to be identified in prediction problems.




A resposta correta é: of the variables that are very different from what is common, and that can hinder the detection of patterns.

### Pergunta 5

Correta Pontuação 1,0 de 1,0 Retirar destaque

Feature selection methods used before modeling:


- ☐ a. Não quero responder
- ☐ b. eliminate only variables with no useful information for the model.
- ☒ c. can eliminate both redundant and variable variables without useful information for the model. 
- ☐ d. eliminate only redundant variables.
- ☐ e. eliminate redundant or variable variables with no useful information for the model, but the same method never eliminates both types of variables.

A resposta correta é: can eliminate both redundant and variable variables without useful information for the model.

### Pergunta 6

Correta Pontuação 1,0 de 1,0 Destacar pergunta

Regarding the data, the main difference between regression and classification is:

- ☐ a. in classification, the independent variable is nominal, while in regression it is numerical.
- ☐ b. in classification, the independent variable is numerical, while in regression it is nominal.
- ☐ c. in classification, the objective variable is numerical, whereas in regression it is nominal.
- ☒ d. in classification, the target variable is nominal, while in regression it is numerical. 

### Pergunta 7

Incorreta Pontuou -0,2 de 1,0  Destacar pergunta

The curse of dimensionality may occur when:

- ☐ a. the number of examples is insufficient relative to the number of variables, increasing the risk of mistaking noise with patterns.
- ☐ b. the number of variables is insufficient compared to the number of examples, increasing the risk of mistaking noise with patterns.
- ☐ c. Não quero responder
- ☒ d. the number of examples insufficient relative to the number of variables, increasing the risk of missing relevant patterns.
- ☐ e. the number of variables is insufficient compared to the number of examples, increasing the risk of missing relevant patterns.



A resposta correta é: the number of examples is insufficient relative to the number of variables, increasing the risk of mistaking noise with patterns.

### Pergunta 8

Incorreta Pontuou 0,0 de 1,0  Retirar destaque

Comparing with the MSE measure (mean squared error), the MAD measure (mean absolute deviation):

- ☐ a. places more emphasis on larger errors.
- ☐ b. places more emphasis on minor errors.
- ☐ c. ignores errors that are too large to be credible.
- ☐ d. treats all errors equally.
- ☒ e. Não quero responder



A resposta correta é: treats all errors equally.

## Pergunta 9

Incorreta

Pontuou -0,2 de 1,0

🚩 Destacar pergunta

One of the most important aspects of developing predictive models is:

- ☐ a. the estimate of its computational performance. For the value obtained to be a reliable measure of the computational cost of the forecast, it is essential that the data is fed to the model with the same speed that would happen in reality and that the sampling is completely random.
- ☐ b. Não quero responder
- ☐ c. the estimate of its predictive performance. For this, it is essential that data other than training and testing are used. But sometimes that is not enough. It is also necessary to ensure that the test data was generated after the training data.
- ☐ d. the estimate of its predictive performance. For the value obtained to be a reliable measure of overfitting capacity, it is essential that different training and test data are used and that sampling is completely random.
- ☒ e. the estimate of its predictive performance. For the value obtained to be a reliable measure of generalizability, it is essential that different training and test data are used and that sampling is completely random. ❌

A resposta correta é: the estimate of its predictive performance. For this, it is essential that data other than training and testing are used. But sometimes that is not enough. It is also necessary to ensure that the test data was generated after the training data.

## Pergunta 10

Correta

Pontuou 1,0 de 1,0

🚩 Retirar destaque

What is the most appropriate phrase to characterize the k-nearest neighbours algorithm:

- ☒ a. Its predictive performance is sensitive to irrelevant attributes. ✓
- ☐ b. Não quero responder
- ☐ c. It is equally prepared to deal with numeric and nominal attributes.
- ☐ d. IT does not have hyperparameters.
- ☐ e. Its predictive performance is robust to redundant attributes.

### Pergunta 11

Correta Pontuou 1,0 de 1,0 Destacar pergunta

What is the most appropriate phrase to characterize the naive bayes algorithm:

- ☐ a. It is equally prepared to deal with numeric and nominal attributes.
- ☒ b. Its predictive performance is robust to irrelevant attributes.
- ☐ c. The tuning of the hyperparameters is complicated.
- ☐ d. Não quero responder
- ☐ e. Its predictive performance is sensitive to redundant attributes.



A resposta correta é: Its predictive performance is robust to irrelevant attributes.

### Pergunta 12

Correta Pontuou 1,0 de 1,0 Destacar pergunta

The backpropagation algorithm in neural networks:

- ☐ a. Não quero responder
- ☒ b. propagates the inputs forward and the errors backwards.
- ☐ c. propagates the inputs forward and the predictions backwards.
- ☐ d. propagates the errors forward and the inputs backwards.
- ☐ e. propagates the predictions forward and the errors backwards.



A resposta correta é: propagates the inputs forward and the errors backwards.

### Pergunta 13

Correta Pontuou 1,0 de 1,0 Destacar pergunta

How do machines with support vectors guarantee good performance in non-linear problems?

- ☐ a. via the epsilon hyperparameter.
- ☐ b. through the use of the soft-margin.
- ☐ c. through the hyper-parameter cost.
- ☐ d. Não quero responder
- ☒ e. through the use of kernels.



A resposta correta é: through the use of kernels.

### Pergunta 14

Correta Pontuou 1,0 de 1,0 Destacar pergunta

In general, ensembles are sets of models that take advantage of:

- ☒ a. of the differences between the areas of the space in which different models are competent.
- ☐ b. Não quero responder
- ☐ c. of the similarities between the areas of the in which different models are competent.
- ☐ d. to repeat the execution of the algorithms on the same data, which allows to obtain better estimates of performance.
- ☐ e. to repeat the execution of the algorithms on slightly different data, which allows to obtain better performance estimates.



A resposta correta é: of the differences between the areas of the space in which different models are competent.



### Pergunta 15

Correta Pontuou 1,0 de 1,0 Destacar pergunta

Which of the following statements is true:

- ☐ a. The purpose of clustering is only to minimize the intra-cluster distance.
- ☐ b. Não quero responder
- ☒ c. The objective of clustering is to maximize the inter-cluster distance and minimize the intra-cluster distance. ✓
- ☐ d. The purpose of clustering is only to minimize the inter-cluster distance.
- ☐ e. The objective of clustering is to maximize the intra-cluster distance and minimize the inter-cluster distance.

A resposta correta é: The objective of clustering is to maximize the inter-cluster distance and minimize the intra-cluster distance.

### Pergunta 16

Incorreta Pontuou -0,2 de 1,0 Destacar pergunta


Indicate which of the following options describes important characteristics of the clusterings obtained with the k-means and DBSCAN algorithms:

- ☐ a. K-means is robust to outliers whereas DBSCAN is sensitive to outliers.
- ☒ b. The centroids of the clusters obtained with both algorithms can be interpreted as profiles of the respective examples. ✗
- ☐ c. Clusters obtained with k-means are spherical whereas those obtained with DBSCAN may have arbitrary forms.
- ☐ d. DBSCAN is computationally more efficient than k-means.
- ☐ e. Não quero responder

A resposta correta é: Clusters obtained with k-means are spherical whereas those obtained with DBSCAN may have arbitrary forms.



### Pergunta 17

Correta Pontuou 1,0 de 1,0  Destacar pergunta

Assuming the set of transactions  $\{\{A, B\}, \{A, B\}, \{A, C\}, \{A, D\}, \{D\}\}$ , indicate which sets are frequent for a minimum support of 0.45:

- ☐ a. Não quero responder
- ☒ b.  $\{A\}$
- ☐ c.  $\{A, B\}$
- ☐ d.  $\{A\}$ ,  $\{B\}$ ,  $\{D\}$  and  $\{A, B\}$
- ☐ e.  $\{A, B\}$  and  $\{A, D\}$



A resposta correta é:  $\{A\}$

### Pergunta 18

Incorreta Pontuou 0,0 de 1,0  Retirar destaque

In association rules, greater confidence is usually associated with:

- ☒ a. Não quero responder
- ☐ b. greater support, as it typically involves less specific rules.
- ☐ c. less support, as it typically involves more specific rules.
- ☐ d. greater support, since it typically involves rules with fewer items in the consequent.
- ☐ e. less support, since it typically involves rules with fewer items in the antecedent.



A resposta correta é: less support, as it typically involves more specific rules.

SMOTE is a technique used when:

- ☐ a. The dataset is for classification.
- ☒ b. The dataset is unbalanced.
- ☐ c. The dataset is for regression.
- ☐ d. The dataset is balanced.
- ☐ e. Não quero responder



A resposta correta é: The dataset is unbalanced.

## Pergunta 20

Correta Pontuação 1,0 de 1,0 Destacar pergunta

In spatial-temporal data analysis, it is assumed that two observations have:

- ☐ a. Não quero responder
- ☐ b. only 1 type of dependency, either temporal (when closer in time, greater dependency) or spatial (when closer in space, greater dependency).
- ☒ c. at least 2 types of dependencies, temporal (when closer in time, greater dependence) and spatial (when closer in space, greater dependence).
- ☐ d. only 1 type of dependency, temporal (when closer in time, greater dependence) but not spatial (distance in space is irrelevant for the analysis).
- ☐ e. only 1 type of dependence, spatial (when closer in space, the greater the dependence) but not temporal (distance in time is irrelevant for the analysis).



A resposta correta é: at least 2 types of dependencies, temporal (when closer in time, greater dependence) and spatial (when closer in space, greater dependence).